

# Notice of Allowability

Application No.

09/918,952

Examiner

JEAN B. FLEURANTIN

Applicant(s)

YU ET AL.

Art Unit

2162

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.


1. ☒ This communication is responsive to 29 September 2004.
2. ☒ The allowed claim(s) is/are 26-34.
3. ☒ The drawings filed on 7/31/01 are accepted by the Examiner.
4. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
  - a) ☐ All b) ☐ Some\* c) ☐ None of the:
    1. ☐ Certified copies of the priority documents have been received.
    2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
    3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).
  - \* Certified copies not received: \_\_\_\_\_.

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.  
**THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.**

5. ☐ A SUBSTITUTE OATH OR DECLARATION must be submitted. Note the attached EXAMINER'S AMENDMENT or NOTICE OF INFORMAL PATENT APPLICATION (PTO-152) which gives reason(s) why the oath or declaration is deficient.
  6. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
    - (a) ☐ including changes required by the Notice of Draftsperson's Patent Drawing Review (PTO-948) attached
      - 1) ☐ hereto or 2) ☐ to Paper No./Mail Date \_\_\_\_\_.
    - (b) ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date \_\_\_\_\_.
- Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).
7. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

## Attachment(s)

1. ☐ Notice of References Cited (PTO-892)
2. ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3. ☐ Information Disclosure Statements (PTO-1449 or PTO/SB/08), Paper No./Mail Date \_\_\_\_\_
4. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material
5. ☐ Notice of Informal Patent Application (PTO-152)
6. ☐ Interview Summary (PTO-413), Paper No./Mail Date \_\_\_\_\_
7. ☐ Examiner's Amendment/Comment
8. ☒ Examiner's Statement of Reasons for Allowance
9. ☐ Other \_\_\_\_\_

  
**SHAHID ALAM**  
**PRIMARY EXAMINER**

### **DETAILED ACTION**

1. This is in response to the Applicant's arguments filed 29 September 2004, with respect to claims 26-34 have been fully considered and are persuasive, in which claims 26-34 remain pending.

### **REASONS FOR ALLOWANCE**

2. With respect to claims 26-34 are allowed over the prior art of record.

The following is an examiner's statement of reasons for allowance:

With respect to claim 26, the claimed features "for a current leaf node among the leaf nodes of the decision tree, computing a lowest value of a gini index achieved by univariate-based partitions on each of a plurality of attribute lists included in the current leaf node; and wherein the gini index is equal to  $1 - (P_n)^2 - (P_p)^2$ ,  $P_n$  being a percentage of the records of the non-target class in the input data set and  $P_p$  being a percentage of the records of the target class in the input data set" in conjunction with other elements of the independent claims would not found anticipated or obvious over the prior art made of record. With respect to claim 27, the claimed features "for a current leaf node from among the leaf nodes of the decision tree, computing a lowest value of a gini index achieved by univariate-based partitions on each of a plurality of attribute lists included in the current leaf node; and wherein the percentage of the records  $P_p$  in the input data set is equal to  $W_p * n_p / (W_p * n_p + n_n)$ ,  $W_p$  being a weight of the records of the target class in the input data set,  $n_p$  and  $n_n$  being a number of the records of the target class and a number of the records of the non-target

Art Unit: 2162

class in the current leaf node, respectively” in conjunction with other elements of the independent claims would not found anticipated or obvious over the prior art made of record. With respect to claim 28, the claimed features “computing a lowest value of a gini index achieved by univariate-based partitions on each of a plurality of attribute lists included in the current leaf node; and wherein said partitioning step further comprises the steps of: detecting subspace clusters of the records of the target class associated with the current leaf node; computing the lowest value of the gini index achieved by distance-based partitions on each of the plurality of attribute lists, included in the current leaf node, the distance based partitions being based on distances to the detected subspace clusters; partitioning pre-sorted attribute lists included in the current node into two sets of ordered attribute lists based upon a greater one of the lowest value of the gini index achieved by univariate partitions and the lowest value of the gini index achieved by distance-based partitions; and creating new child nodes for each of the two sets of ordered attribute lists; and wherein said detecting step comprises the steps of: computing a minimum support (minsup) of each of the subspace clusters that have a potential of providing a lower gini index than that provided by the univariate-based partitions; identifying one-dimensional clusters of the records of the target class associated with the current leaf node; beginning with the one-dimensional clusters, combining centroids of K dimensional clusters to form candidate (K+1)-dimensional clusters; identifying a number of the records of the target class that fall into each of the (K+1)-dimensional clusters; pruning any of the (K+1)-dimensional clusters that have a support lower than the minsup” in conjunction with other elements of the independent

Art Unit: 2162

claims would not found anticipated or obvious over the prior art made of record. With respect to claim 32, the claimed features "detecting subspace clusters of the records of the target class associated with the current leaf node; computing the lowest value of the gini index achieved by distance-based partitions on each of the plurality of attribute lists included in the current leaf node, the distance based partitions being based on distances to the detected subspace clusters; partitioning pre-sorted attribute lists included in the current node into two sets of ordered attribute lists based upon a greater one of the lowest value of the gini index achieved by univariate partitions and the lowest value of the gini index achieved by distance-based partitions; and creating new child nodes for each of the two sets of ordered attribute lists; and wherein said step of computing the lowest value of the gini index achieved by distance-based partitions comprises the steps of: identifying eligible subspace clusters from among the subspace clusters, an eligible subspace cluster having a set of clustered dimensions such that only less than all of the clustered dimensions in the set are capable of being included in another set of clustered dimensions of another subspace cluster; selecting top-K clusters from among the eligible subspace clusters, the top-K clusters being ordered by a number of records therein; for each of a current top-K cluster, computing a centroid of the current top-K cluster and a weight on each dimension of the current top-K cluster; and computing the gini index of the current top-K cluster, based on a weighted Euclidean distance to the centroid; and recording a lowest gini index achieved by said step of computing the gini index of the current top-K cluster" in conjunction with other elements of the independent claims would not found anticipated or obvious over the

Art Unit: 2162

prior art made of record. With respect to claim 33, the claimed features “computing the lowest value of the gini index achieved by distance-based partitions on each of the plurality of attribute lists included in the current leaf node, the distance based partitions being based on distances to the detected subspace clusters; partitioning pre-sorted attribute lists included in the current node into two sets of ordered attribute lists based upon a greater one of the lowest value of the gini index achieved by univariate partitions and the lowest value of the gini index achieved by distance-based partitions; and creating new child nodes for each of the two sets of ordered attribute lists; and wherein each of the plurality of pre-sorted attribute lists comprises a plurality of entries, and said step of partitioning the pre-sorted attribute lists comprises the steps of determining whether univariate partitioning or distance-based partitioning has occurred; creating a first hash table that maps record ids of any of the records that satisfy a condition  $A=v$  to a left child node and that maps the record ids of any of the records that do not satisfy the condition  $A=v$  to a right child node,  $A$  being an attribute and  $v$  denoting a splitting position, when the univariate partitioning has occurred; creating a second hash table that maps the record ids of any of the records that satisfy a condition  $\text{Dist}(d, p, w)=v$  to a left child node and that maps the record ids of any of the records that do not satisfy the condition  $\text{Dist}(d, p, w)=v$  to a right child node, when the distance-based partitioning has occurred,  $d$  being a record associated with a current subspace cluster,  $p$  being a centroid of the current subspace cluster, and  $w$  being a weight on dimensions of the current subspace cluster; partitioning the pre-sorted attribute lists into the two sets of ordered attribute lists, based on information in a corresponding one of the first hash

Art Unit: 2162

table or the second hash table; appending each entry of the two sets of ordered attribute lists to one of the left child node or the right child node, based on the information in the corresponding one of the first hash table or the second hash table and information corresponding to the each entry, to maintain attribute ordering in the two sets of ordered attribute lists that corresponds that in the pre-sorted attribute lists" in conjunction with other elements of the independent claims would not found anticipated or obvious over the prior art made of record. With respect to claim 34, the claimed features "wherein said classifying and scoring step comprises the steps of for each of the plurality of nodes of the decision tree, starting at the root node, evaluating a Boolean condition and following at least one branch of the decision tree until a leaf node is reached; classifying the reached leaf node based on a majority class of any of the predetermined attributes included therein; for each node in the nearest neighbor set of nodes for the reached leaf node, computing a distance between a record to be scored and a centroid of the reached leaf node, using a distance function computed for the reached leaf node; and scoring the record using a maximum value of a score function, the score function defined as  $\text{conf}/\text{dist}(d,p,w,)$ , wherein the conf is a confidence of the reached node, d is a particular record associated with a current subspace cluster, p is a centroid of the current subspace cluster, and w is a weight on dimensions of the subspace cluster " in conjunction with other elements of the independent claims would not found anticipated or obvious over the prior art made of record.

The dependent claims, being definite, further limiting, and fully enabled by the specification are also allowed.

Art Unit: 2162

3. The closest prior art, Agrawal et al. U.S. Patent No. 5,799,311 relates to data mining. Ramaswamy et al. "Efficient Algorithms for Mining Outliers from Large Data Sets" but fail to teach the above limitations.

4. Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

### CONTACT INFORMATION


5. Any inquiry concerning this communication or earlier communications from the examiner should be directed to JEAN B. FLEURANTIN whose telephone number is 571 - 272-4035. The examiner can normally be reached on 7:05 to 4:35.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, JOHN E BREENE can be reached on 571 - 272-4107. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

  
Jean Bolte Fleurantin

October 29, 2004

  
SHAHID ALAM  
PRIMARY EXAMINER